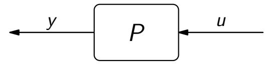
Introduction to Control (00340040)
lecture no. 3

Leonid Mirkin

Faculty of Mechanical Engineering
Technion—IIT

Abstract control problem to begin with (contd)

Setup:



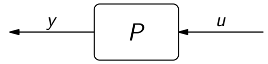
where

– P is a plant
  may comprise actual controlled process, actuators, sensors, et cetera
– u is a control signal (control input)
– y is a controlled (regulated) signal (output)

Problem: Given P, find u resulting in a desired y.

Outline

Desired behavior



"Desired y" may be introduced via the requirement that

$$y = r$$

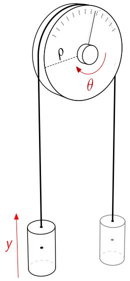
for some signal

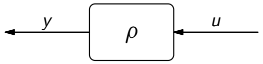
– r, called the reference trajectory.

Reference trajectories can be produced

– offline $\implies$ available in whole during the operation
  e.g. elevator / crane / printer hear position profiles
– online $\implies$ retrieved from measured data
  e.g. cable-suspended cameras in sport events / missile interceptors

Control of Σ_1



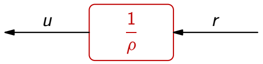


Our goal:

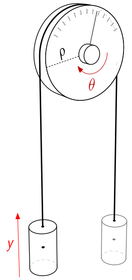
– find u such that $y = r$ for a given signal r.

Obviously,

$$y = \rho u \quad \wedge \quad y = r$$
$$\Updownarrow$$
$$r = \rho u$$
$$\Updownarrow$$
$$u = \frac{1}{\rho} r \iff$$

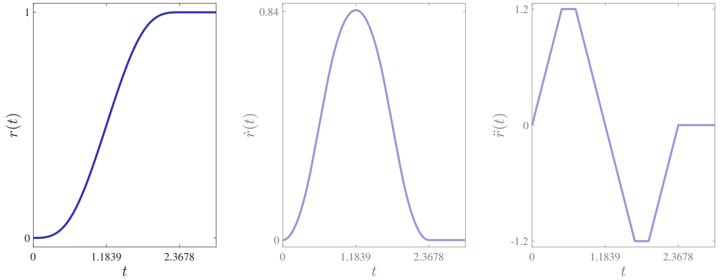


Control of Σ_1: example



$\rho = 1.25\,\text{m}$

The choice of the reference signal:



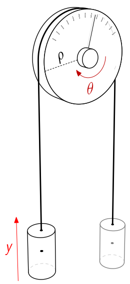
Remarks:

– both $\dot r$ and $\ddot r$ are bounded and continuous,

– the fastest rise by $1\,[\text{m}]$ subject to $|\ddot r(t)| \le 1.2\,[\text{m/s}^2]$ and $|\dddot r(t)| \le 2.5\,[\text{m/s}^3]$ for all $t \ge 0$,

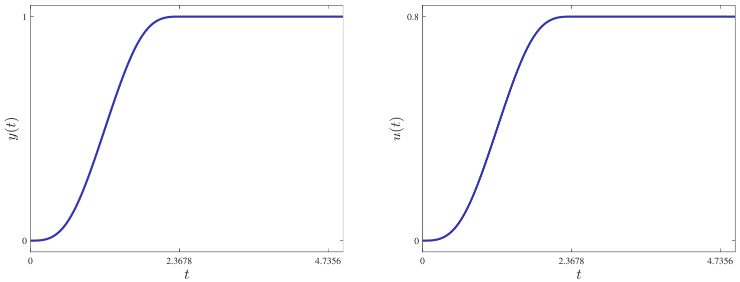
– the final position, viz. $\lim_{t\to\infty} y(t) = 1\,[\text{m}]$, is selected to emphasize transition regimes.

Control of Σ_1: example (contd)



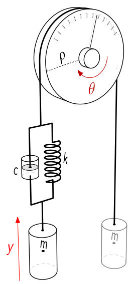
$\rho = 1.25\,\text{m}$

Applying $u = 0.8r$,



This control trajectory is easy to guess.

Control of Σ_2: example



$\rho = 1.25\,\text{m}$

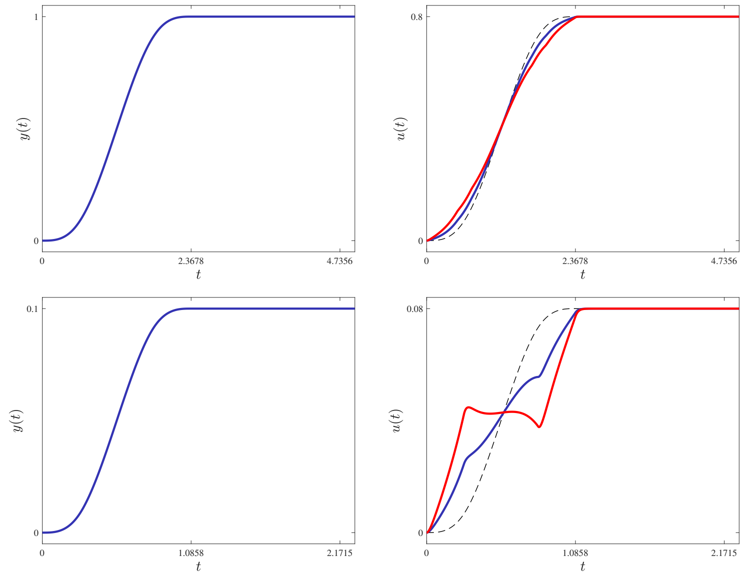
$k = 40000\,\frac{\text{N sec}}{\text{m}}$

$c = 800\,\frac{\text{N}}{\text{m}}$

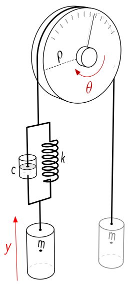
$m = 1410\,\text{kg}$

$m = 2820\,\text{kg}$

A guess would be harder for a more complex plant:



Control of Σ_2



$$y \longleftarrow \boxed{\frac{(cs+k)\rho}{ms^2+cs+k}} \overset{u}{\longleftarrow}$$

With the same goal as above,

$$m\ddot{y} + c\dot{y} + ky = \rho(c\dot{u} + ku) \quad \wedge \quad y = r$$
$$\Updownarrow$$
$$\rho(c\dot{u} + ku) = m\ddot{r} + c\dot{r} + kr$$
$$\Updownarrow$$
$$u \longleftarrow \boxed{\frac{ms^2+cs+k}{(cs+k)\rho}} \overset{r}{\longleftarrow}$$

(as $\rho(c\dot{u} + ku) = m\ddot{r} + c\dot{r} + kr \iff U(s) = \frac{ms^2+cs+k}{(cs+k)\rho}R(s)$).

Open-loop control

Control systems above can be visualized as

$$\overset{y}{\longleftarrow} \boxed{\rho} \overset{u}{\longleftarrow} \boxed{\frac{1}{\rho}} \overset{r}{\longleftarrow}$$

and

$$\overset{y}{\longleftarrow} \boxed{\frac{(cs+k)\rho}{ms^2+cs+k}} \overset{u}{\longleftarrow} \boxed{\frac{ms^2+cs+k}{(cs+k)\rho}} \overset{r}{\longleftarrow}$$

They are both particular cases of the open-loop control scheme

$$\overset{y}{\longleftarrow} \boxed{P} \overset{u}{\longleftarrow} \boxed{C_{ol}} \overset{r}{\longleftarrow}$$

where C_{ol} is the controller connected in series with the plant and generating the control signal $u = C_{ol}r$.

Plant inversion

Moreover, controllers in these schemes,

$$\overset{y}{\longleftarrow} \boxed{\rho} \overset{u}{\longleftarrow} \boxed{\frac{1}{\rho}} \overset{r}{\longleftarrow}$$

$$\overset{y}{\longleftarrow} \boxed{\frac{(cs+k)\rho}{ms^2+cs+k}} \overset{u}{\longleftarrow} \boxed{\frac{ms^2+cs+k}{(cs+k)\rho}} \overset{r}{\longleftarrow}$$

act according to the same principle:

$$\overset{y}{\longleftarrow} \boxed{P} \overset{u}{\longleftarrow} \boxed{P^{-1}} \overset{r}{\longleftarrow}$$

i.e. the controller

$$C_{ol} = P^{-1},$$

where P^{-1} is the system such that $y = Pu \implies u = P^{-1}y$, whose transfer function equals $1/P(s)$. This strategy is called plant inversion.

Plant inversion (contd)

As straightforward as it may look, this idea is

- the heart of most (model-based) control strategies.

The controller

$$\overset{y}{\longleftarrow} \boxed{P} \overset{u}{\longleftarrow} \boxed{P^{-1}} \overset{r}{\longleftarrow}$$

guarantees $y = r$ for every r.

Q: Is it *that* simple?

A: No, it is not.

Outline

Open-loop control

Plant inversion: some limitations

Limitations of plant inversion: internal stability

Steady-state and transient performance

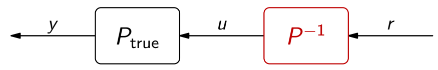
Transient responses of 1st and 2nd order systems (self-study, from LS)

Modeling uncertainty

Remember,

– models of real-world phenomena are never perfect.

Thus, P in $C_{\mathrm{ol}} = P^{-1}$ is not the plant, but rather its (more or less accurate) approximation and the actual controlled system looks more like



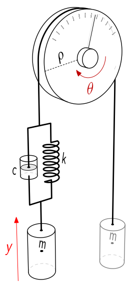
for some "true" plant P_{true}. We then have that

$$y = P_{\mathrm{true}} P^{-1} r$$

and this $y \neq r$ whenever $P_{\mathrm{true}} \neq P$, with the error

$$e := r - y = r - P_{\mathrm{true}} P^{-1} r = \underbrace{(1 - P_{\mathrm{true}} P^{-1})}_{\Delta_P} r$$

Modeling uncertainty & plant inversion: example



$\rho = 1.25\,\mathrm{m}$

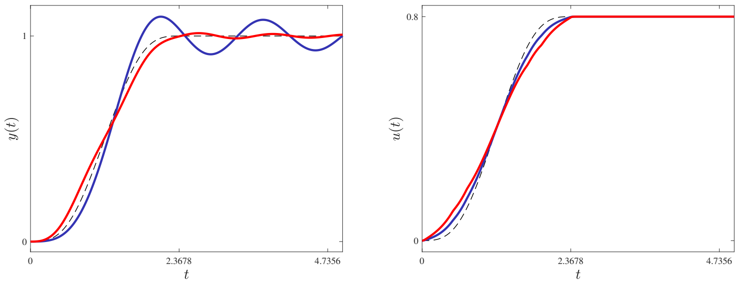
$k = 40000\,\frac{\mathrm{N\,sec}}{\mathrm{m}}$

$c = 800\,\frac{\mathrm{N}}{\mathrm{m}}$

$m = 1410\,\mathrm{kg}$

$m = 2820\,\mathrm{kg}$

If the actual mass mismatches that assumed in the design of u:



– u calculated for $m = 1410$ applied to $m = 2820$

– u calculated for $m = 2820$ applied to $m = 1410$

Curiously,

– "blue" oscillations are substantially larger than "red" ones.

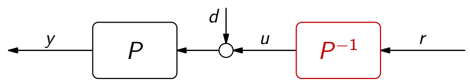
Why?

Disturbances

Controlled systems

– always interact with the environment.

A way to express such interactions is by introducing disturbances, which are exogenous signals (i.e. independent of control actions) affecting the system. An example is the load disturbance d acting at the input and leading to



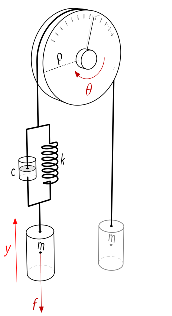
In this case

$$y = P(P^{-1} r + d) = r + Pd$$

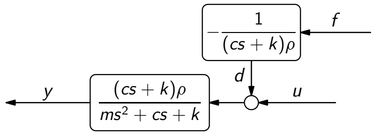
and this $y \neq r$ whenever $Pd \neq 0$, with the error

$$e = r - y = -Pd$$

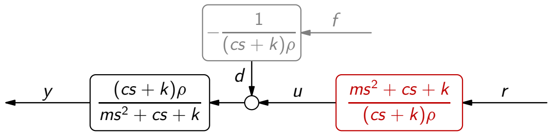
Disturbances & plant inversion: example



The effect of external force f corresponds to the diagram

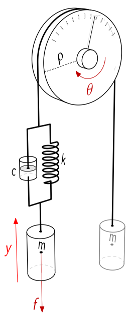


i.e. this d is a low-pass filtered and scaled (by $\frac{1}{k\rho}$) version of f. Open-loop controlled system is then



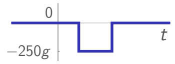
$\rho = 1.25\,\mathrm{m}$
$k = 40000\,\frac{\mathrm{N\,sec}}{\mathrm{m}}$
$c = 800\,\frac{\mathrm{N}}{\mathrm{m}}$
$m = 1410\,\mathrm{kg}$
$m = 2820\,\mathrm{kg}$

Disturbances & plant inversion: example (contd)

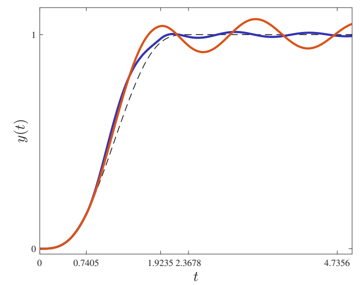


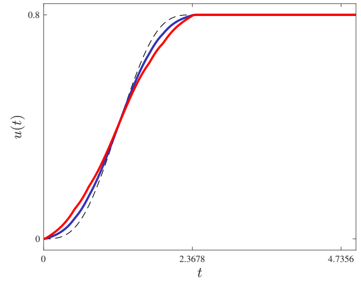
Let[a]

$$f = -250g(\mathbb{S}_{-0.74}\mathbb{1} - \mathbb{S}_{-1.92}\mathbb{1}) =$$



(think of a jump of somebody heavy, like , in an elevator). Responses:



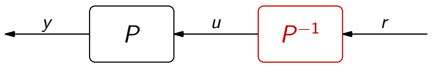


Now "red" oscillations are substantially larger than "blue" ones. Why?

$\rho = 1.25\,\mathrm{m}$
$k = 40000\,\frac{\mathrm{N\,sec}}{\mathrm{m}}$
$c = 800\,\frac{\mathrm{N}}{\mathrm{m}}$
$m = 1410\,\mathrm{kg}$
$m = 2820\,\mathrm{kg}$

[a] The time shift operator $\mathbb{S}_\tau$ acts as $(\mathbb{S}_\tau x)(t) = x(t+\tau)$.

Control effort



The control signal

$$u = P^{-1}r$$

is fed to an actuator. Every physical actuator has limitations, like

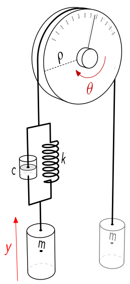
- bounded input amplitude
- bounded input rate
- ...

We also prefer "smaller" and "smoother" control signals because of other considerations (like energy consumption, equipment wear and tear, etc)

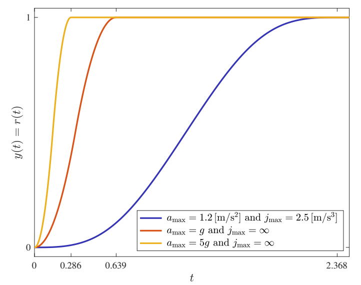
Dynamic relations between r and u might result in

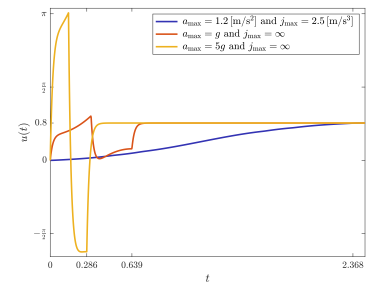
- unacceptable control signals from seemingly innocent references.

Control effort & plant inversion: example



Fastest rise by 1[m] under $|\ddot{r}(t)| \le a_{\max}$ and $|\dddot{r}(t)| \le j_{\max}$ for various $a_{\max}$ and $j_{\max}$:

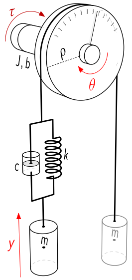




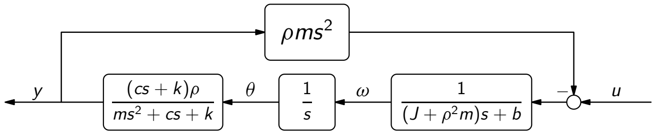
This might require high control effort (large amplitude of u). Why?

$\rho = 1.25\,\mathrm{m}$
$k = 40000\,\frac{\mathrm{N\,sec}}{\mathrm{m}}$
$c = 800\,\frac{\mathrm{N}}{\mathrm{m}}$
$m = 2820\,\mathrm{kg}$

Control of Σ_3



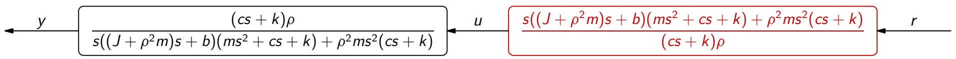
With $u = \tau$ (torque),



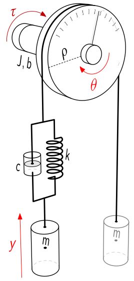
for which

$$P(s) = \frac{(cs+k)\rho}{s((J+\rho^2 m)s+b)(ms^2+cs+k)+\rho^2 m s^2(cs+k)}$$

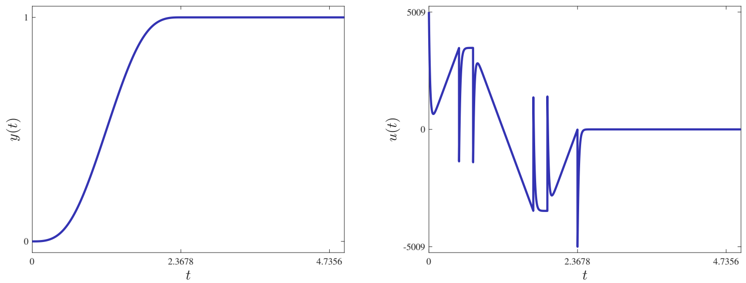
Plant inversion works then as follows:



Control of Σ_3: example



We still have perfect response



requiring sophisticated control trajectory

$\rho = 1.25\,\mathrm{m}$

$k = 40000\,\frac{\mathrm{N\,sec}}{\mathrm{m}}$

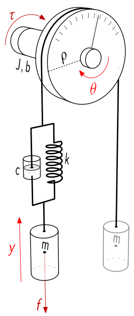
$c = 800\,\frac{\mathrm{N}}{\mathrm{m}}$

$m = 1410\,\mathrm{kg}$

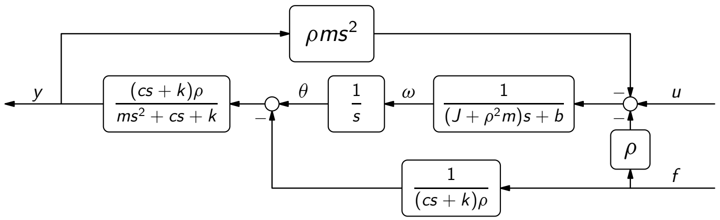
$J = 11\,\mathrm{kg\,m^2}$

$b = 0$

Control of Σ_3: adding disturbances



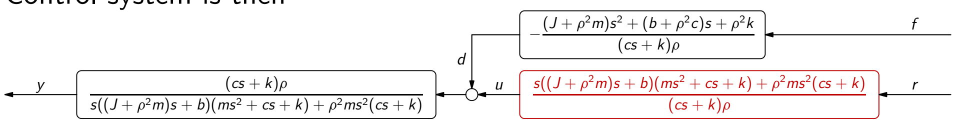
If an external force f applies to the mass, we have:



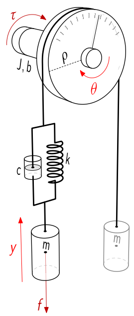
which is equivalent to applying input disturbance d such that

$$D(s) = -\frac{(J+\rho^2 m)s^2+(b+\rho^2 c)s+\rho^2 k}{(cs+k)\rho}\,F(s)$$

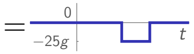
Control system is then



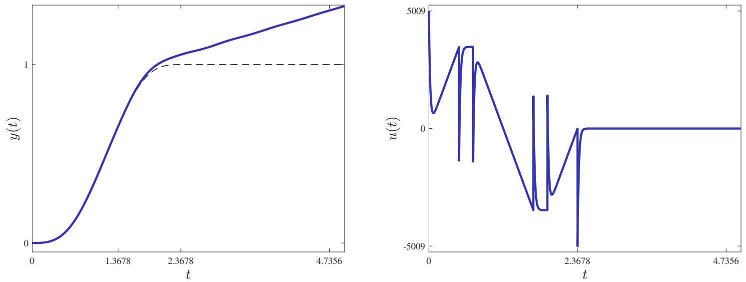
Control of Σ_3: example (contd)



If

$$f = -25g(\mathbb{S}_{-1.3678}\mathbb{1} - \mathbb{S}_{-2.3678}\mathbb{1}) =$$ 

(like a jump of somebody light, like , in an elevator), then



$\rho = 1.25\,\mathrm{m}$

$k = 40000\,\frac{\mathrm{N\,sec}}{\mathrm{m}}$

$c = 800\,\frac{\mathrm{N}}{\mathrm{m}}$

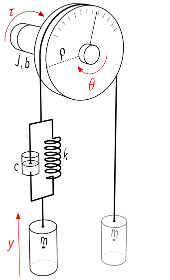
$m = 1410\,\mathrm{kg}$

$J = 11\,\mathrm{kg\,m^2}$

$b = 0$

Oops... Explanations?

Control of Σ_3: example (contd)



$\rho = 1.25\,\mathrm{m}$

$k = 40000\,\frac{\mathrm{N\,sec}}{\mathrm{m}}$

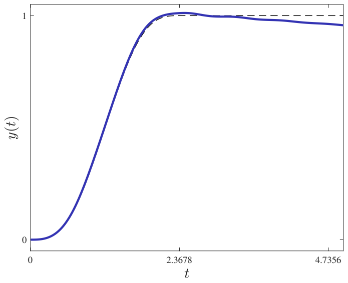
$c = 800\,\frac{\mathrm{N}}{\mathrm{m}}$

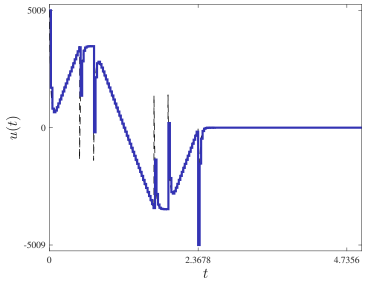
$m = 1410\,\mathrm{kg}$

$J = 11\,\mathrm{kg\,m^2}$

$b = 0$

Let the controller be implemented digitally, so that the control trajectory is piecewise-constant:





Now, seemingly small deviations of the control signal from its designed waveform yields a steady drift of the regulated signal away from the required value. Explanations ?!!

Outline

Open-loop control

Plant inversion: some limitations

Limitations of plant inversion: internal stability

Steady-state and transient performance

Transient responses of 1st and 2nd order systems (self-study, from LS)

Preliminaries: BIBO stability (from LS)

A linear $G : u \mapsto y$ is said to be BIBO stable if

- $\exists \gamma \geq 0$, independent of u, such that $\|y\|_\infty \leq \gamma \|u\|_\infty$ for all $u \in L_\infty$

(i.e. a bounded input always results in a bounded output, hence the name). If G_1 and G_2 are stable, then so are $G_2 G_1$ and $G_2 + G_1$.

Remember (from LS) that

$$L_q := \{x : \mathbb{R} \to \mathbb{R} \mid \|x\|_q < \infty\}, \text{ where } \|x\|_\infty := \sup_{t\in\mathbb{R}} |x(t)| \text{ and } \|x\|_1 := \int_{\mathbb{R}} |x(t)|\mathrm{d}t.$$

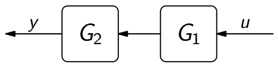
If G is LTI, then

- G is BIBO stable iff its impulse response $g \in L_1$.

If G is LTI and its transfer function $G(s)$ is rational, i.e. $G(s) = \dfrac{N_G(s)}{D_G(s)}$ for polynomials $N_G(s)$ and $D_G(s)$, then it is BIBO stable iff

1. $G(s)$ is proper ($\deg D_G(s) \geq \deg N_G(s)$)
2. $G(s)$ has no poles in the closed RHP $\bar{\mathbb{C}}_0 := \{s \in \mathbb{C} \mid \operatorname{Re} s \geq 0\}$

Preliminaries: pole/zero cancellations



We say that there is a cancellation in the series interconnection above if

- the order of the mapping $u \mapsto y$ is smaller than the sum of the orders of its components, G_1 and G_2.

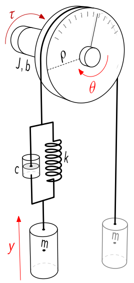
Cancellations mean that some dynamics of the components disappear from the mapping $u \mapsto y$. If disappearing dynamics are unstable, i.e. their pole(s) is in $\bar{\mathbb{C}}_0$, we say that the cancellation is unstable. In the SISO case poles of one component can only be canceled by zeros of the other, hence the term pole/zero cancellations.

Example: Let

$$G_1(s) = \frac{1}{s-a} \quad \text{and} \quad G_2(s) = \frac{s-a}{s+1} \quad \Longrightarrow \quad G_2(s)G_1(s) = \frac{1}{s+1}$$

i.e. the pole of $G_2(s)$ at a is canceled by a zero of $G_2(s)$ (unstable if $a \geq 0$).

Control of Σ_3: cancellations



$b=0$

The plant

$$P(s)=\frac{(cs+k)\rho}{s^2\big((J+\rho^2 m)(ms^2+cs+k)+\rho^2 m(cs+k)\big)}$$

is unstable because of a double pole at the origin. But

$$C_{\rm ol}(s)=\frac{s^2\big((J+\rho^2 m)(ms^2+cs+k)+\rho^2 m(cs+k)\big)}{(cs+k)\rho}$$

cancels all poles and zeros of $P(s)$, so the controlled system $T_{yr}=PC_{\rm ol}:r\mapsto y$ has

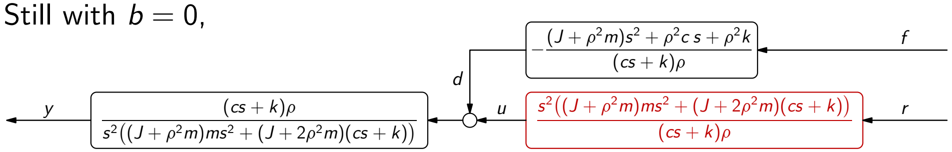
$$T_{yr}(s)=1$$

and is obviously stable. But

– are those cancellations innocent?

Control of Σ_3: cancellations and disturbances

Still with $b=0$,

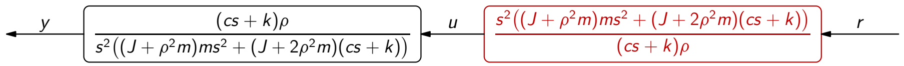


and the system $T_{yf}:f\mapsto y$ has the transfer function

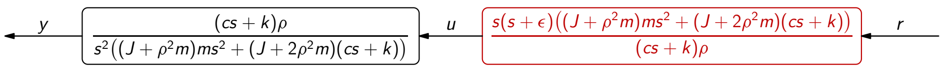
$$T_{yf}(s)=-\frac{(J+\rho^2 m)s^2+\rho^2 c\,s+\rho^2 k}{s^2\big((J+\rho^2 m)ms^2+(J+2\rho^2 m)(cs+k)\big)}.$$

with a double pole at the origin. Hence, T_{yf} is unstable, which explains the problem that we had with the disturbance response.

Control of Σ_3: cancellations and implementation accuracy



yields $T_{yr}(s)=1$ only if the controller is implemented without any errors. If even a small inaccuracy occurs, then the result is different. For example, let the actual controller be



for some ϵ. In this case

$$T_{yr}(s)=\frac{s+\epsilon}{s}$$

is unstable whenever $\epsilon\neq 0$, regardless its size. In other words,

– implementation errors might prevent intended cancellations to happen,

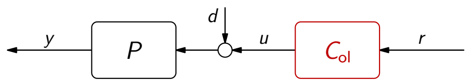
keeping remnants of unwanted plant dynamics in T_{yr}. This explains what we had with the digital implementation.

Internal stability of interconnected systems

The notion of internal stability aims at accounting for implications of

– effects of all exogenous signals on all internal signals

in interconnected systems. Applying to general open-loop control



we consider all mappings between inputs r and d and outputs y and u, i.e.

$$\begin{bmatrix} y \\ u \end{bmatrix}=\begin{bmatrix} PC_{\rm ol} & P \\ C_{\rm ol} & 0 \end{bmatrix}\begin{bmatrix} r \\ d \end{bmatrix}.$$

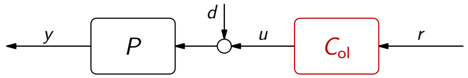
The open-loop control system is said to be

– internally stable if P, $C_{\rm ol}$, and $PC_{\rm ol}$ are all stable.

Because the stability of P and $C_{\rm ol}$ implies that of $PC_{\rm ol}$,

– the system is internally stable iff both P and $C_{\rm ol}$ are stable.

Internal stability and unstable cancellations



By considering

$$\begin{bmatrix} y \\ u \end{bmatrix} = \begin{bmatrix} PC_{\mathrm{ol}} & P \\ C_{\mathrm{ol}} & 0 \end{bmatrix} \begin{bmatrix} r \\ d \end{bmatrix}$$

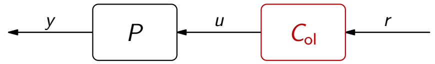
rather than PC_{ol} only, the requirement of internal stability aims at making

– unstable cancellations between P and C_{ol} illegal,

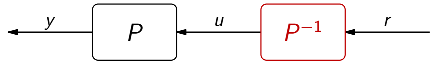
because canceled dynamics of PC_{ol} are still those of P or C_{ol}. Indeed,

– a pole of $P(s)$ canceled by a zero of $C_{\mathrm{ol}}(s)$ is still in $P : d \mapsto y$,
– a pole of $C_{\mathrm{ol}}(s)$ canceled by a zero of $P(s)$ is still in $C_{\mathrm{ol}} : r \mapsto u$.

Internal stability: implications on open-loop control



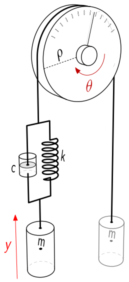
1. not applicable to unstable plants
   no C_{ol} can internally stabilize such systems anyway



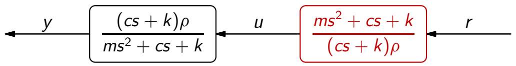
2. not applicable to[1] nonminimum-phase plants
   would result in a controller $C_{\mathrm{ol}} = P^{-1}$ with pole(s) in $\bar{\mathbb{C}}_0$, so unstable
3. might not be applicable to plants with strictly proper transfer functions
   would result in a controller with a non-proper transfer function, so unstable

[1] We say that G is nonminimum-phase if $G(s)$ has at least one zero in $\bar{\mathbb{C}}_0$.

Control of Σ_2: properness



If $c \neq 0$,



then

$$\begin{aligned} U(s) &= \frac{ms^2+cs+k}{(cs+k)\rho} R(s) \\ &= \frac{m}{c\rho} sR(s) + \underbrace{\frac{(c-km/c)s+k}{(cs+k)\rho}}_{\text{stable}} R(s) \end{aligned}$$

This control law implementable only if

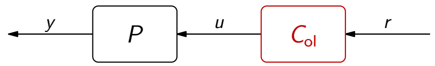
– $\dot{r}$ is bounded and measurable

This might be feasible (esp. if we know the waveform of r in advance), but

– hard (if not impossible) to implement if r is obtained online,

thought a measurement channel.

Controller properness



If $P(s) = N_P(s)/D_P(s)$ with $\deg D_P(s) =: n > m := \deg N_P(s)$, then

$$C_{\mathrm{ol}}(s) = \frac{1}{P(s)} = \frac{D_P(s)}{N_P(s)} = c_{n-m}s^{n-m} + \cdots + c_1 s + \frac{\tilde{D}_P(s)}{N_P(s)}$$

for certain coefficients c_i with $c_{n-m} \neq 0$ and a polynomial $\tilde{D}_P(s)$ such that $\deg \tilde{D}_P(s) = m$. Hence, this C_{ol} is stable and implementable iff

– $N_P(s)$ is Hurwitz and
– $n - m$ derivatives of r are measurable and bounded.

This may be the case if r is generated analytically, by us, but rarely so if r is obtained via sensing a priori unknown signals.

Conclusions

Perfect control, $y = r$,

- is never attainable in practical situations
  because of uncertainty, like modeling errors and disturbances
- might be illegal
  because of internal instability caused by unstable cancellations
- might be too expensive
  in terms of control efforts, reasons not explained yet

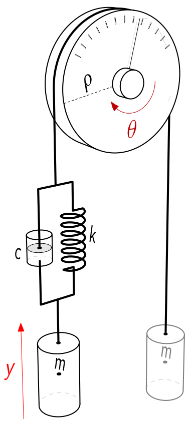
We then shall

- resort to approximate attainment of a desired y, i.e. $y \approx r$
- for a limited class of reference signals r

What could be the meaning of those approximate relation and limited class?

Outline

Requirements on y revisited



With the elevator interpretation in mind, we may dream up requirements like:

1. "move to a given position and stop there," which is our ultimate goal, *where* do we need to go
2. "move fast / slow / smooth / etc," which reflects our anticipations of *how* requirement 1 is met

The control terminology for such requirements is

1. steady-state requirements
2. transient requirements

applicable in various kinds of problems / contexts.

Steady-state and transient responses

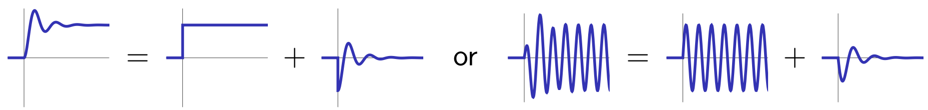
The response of a *stable* LTI system to a regular (e.g. constant, polynomial, periodic) persistent (i.e. non-decaying) input signal u can be decomposed as

$$y = y_{\text{ss}} + y_{\text{tr}},$$

where the steady-state component, y_{ss}, has the same[2] "regularity" as u and the transient component, y_{tr}, decays, i.e. satisfies

$$\lim_{t \to \infty} y_{\text{tr}}(t) = 0.$$

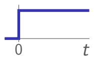
Examples:

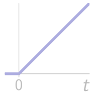


[2] For example, if u is constant (or periodic), then so is y_{ss}.

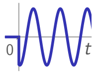
Regular signals that we use

We study mainly responses to polynomial regular signals, like

step: $r(t) = \mathbb{1}(t) =$  with $R(s) = \dfrac{1}{s}$

ramp: $r(t) = \text{ramp}(t) = (\mathbb{1} * \mathbb{1})(t) =$  with $R(s) = \dfrac{1}{s^2}$

and

sine wave: $r(t) = \sin(\omega t + \phi)\mathbb{1}(t) =$  with $R(s) = \dfrac{s \sin\phi + \omega\cos\phi}{s^2 + \omega^2}$

Steady-state specifications: polynomial signals

For polynomial regular signals, like step or ramp, we have clear measure of steady-state performance,

– steady-state error e_{ss}

defined as

$$e_{ss} := \lim_{t\to\infty} |r(t) - y(t)|,$$

where y is the controlled signal. If the controlled system is stable, then the steady-state error can be calculated by the final value theorem. Indeed,

$$y = T_{yr} r \quad \implies \quad e := r - y = (1 - T_{yr}) r$$

If T_{yr} is stable, then

$$e_{ss} = |\lim_{s\to 0} s(1 - T_{yr}(s))R(s)|.$$

and if r is

step: $e_{ss} = |1 - T_{yr}(0)|$ and $T_{yr}(0)$ called static gain of T_{yr}

ramp: $e_{ss} = \left|\lim_{s\to 0} \frac{1 - T_{yr}(s)}{s}\right| = |T'_{yr}(0)|$ (provided $T_{yr}(0) = 1$)

Steady-state specifications: harmonic signals

If $r(t) = \sin(\omega t + \phi)\mathbb{1}(t)$, then the relation

$$r - y = (1 - T_{yr})r$$

and the frequency-response theorem yield

$$r_{ss}(t) - y_{ss}(t) = |1 - T_{yr}(j\omega)| \sin(\omega t + \phi + \arg(1 - T_{yr}(j\omega)))$$

whenever T_{yr} is stable. It is natural to choose

$$e_{ss} = |1 - T_{yr}(j\omega)| = \max_{t\in[0,2\pi/\omega]} |r_{ss}(t) - y_{ss}(t)|$$

as the measure of steady-state mismatch between r and y in this case then (i.e. this e_{ss} is again the steady-state error).

Remark: Mind that $|1 - T_{yr}(j\omega)| \ll 1 \iff |T_{yr}(j\omega)| \approx 1 \wedge \arg T_{yr}(j\omega) \approx 0$.

Steady-state specifications: moral

If the system is stable, then the steady-state error

$$e_{ss} = \begin{cases} |1 - T_{yr}(j\omega)| & \text{if } r(t) = \sin(\omega t + \phi)\mathbb{1}(t) \\ |T'_{yr}(0)| & \text{if } r(t) = \text{ramp}(t) \text{ and } T_{yr}(0) = 1 \end{cases}$$

(step corresponds to $\omega = 0$ and $\phi = \pi/2$) depends only on the

– value of the transfer function $T_{yr}(s)$ at one point at the imaginary axis

or, equivalently, the frequency response $T_{yr}(j\omega)$ at one frequency.

Transient specifications

We prefer transients to be short and smooth. These requirements are often

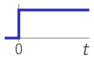
- intrinsically conflicting
- hard to quantify
- heavily dependent on r

It is convenient (e.g. easier to compare, easier to analyze) to

- define transient specifications in terms of the response to a fixed signal

even if the system will actually experience different input signals. Such fixed (test) signal in control is usually the unit step:

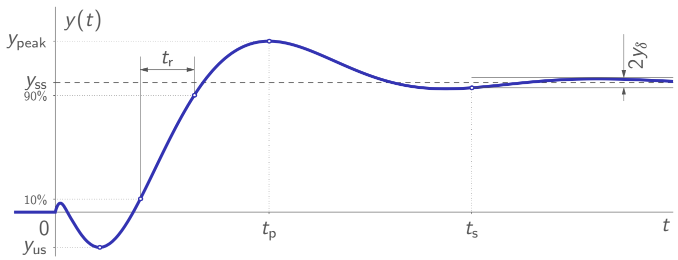
$$\mathbb{1}(t) =$$



because it is

- (relatively) easy to analyze
- "shaky" enough to reveal properties of systems

Step response transient specifications



$\text{OS} := \frac{y_{os}}{y_{ss}}$ overshoot (in %) t_r rise time t_s settling time

$\text{US} := \frac{-y_{us}}{y_{ss}}$ undershoot (in %) t_p peak time $\delta := \frac{y_\delta}{y_{ss}}$ settling level (in %)

- OS and, sometimes, US reflect the "shakiness" of transients
- t_r and, sometimes, t_p reflect the speed of transients
- t_s reflects the duration of transients (given the "duration criterion" δ)

Outline

1st order systems

General form:

$$G(s) = \frac{k_{st}}{\tau s + 1}$$

where

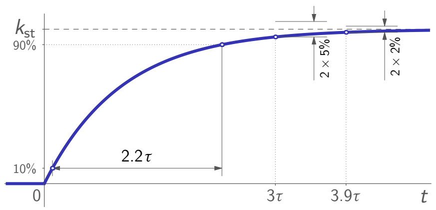
k_{st} static gain

τ time constant (we assume that $\tau > 0$)

Step response:

$$y(t) = k_{st}(1 - e^{-t/\tau})$$

or



with OS = 0%.

2nd order systems

General form:

$$G(s) = \frac{k_{\mathrm{st}}\,\omega_{\mathrm{n}}^2}{s^2 + 2\zeta\omega_{\mathrm{n}} s + \omega_{\mathrm{n}}^2}$$

where

- k_{st} static gain
- ω_{n} natural frequency (we assume that $\omega_{\mathrm{n}} > 0$)
- ζ damping factor (we assume that $\zeta \geq 0$)

This is a second-order system with no zeros.

Classification:

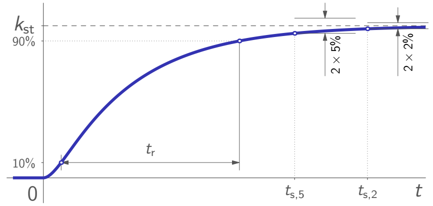
- if $\zeta > 1$ system called overdamped (two distinct real poles)
- if $\zeta = 1$ system called critically damped (double real pole)
- if $\zeta < 1$ system called underdamped (two complex conjugate poles)

2nd order critically and overdamped systems

Step response

$$y(t) = k_{\mathrm{st}}\left(1 - \frac{\zeta + \sqrt{\zeta^2-1}}{2\sqrt{\zeta^2-1}}\,\mathrm{e}^{-(\zeta-\sqrt{\zeta^2-1})\omega_{\mathrm{n}} t} + \frac{\zeta - \sqrt{\zeta^2-1}}{2\sqrt{\zeta^2-1}}\,\mathrm{e}^{-(\zeta+\sqrt{\zeta^2-1})\omega_{\mathrm{n}} t}\right)$$

(for critically damped systems, $y(t) = k_{\mathrm{st}}(1 - \mathrm{e}^{-\omega_{\mathrm{n}} t} - \omega_{\mathrm{n}} t\,\mathrm{e}^{-\omega_{\mathrm{n}} t})\mathbb{1}(t)$) or



$$t_{\mathrm{r}} \approx \frac{1}{\omega_{\mathrm{n}}}\begin{cases} 4.9\zeta - 1.5 & \text{if } 1 \leq \zeta < 2 \\ 4.4\zeta - 0.5 & \text{if } \zeta \geq 2 \end{cases}$$

$$t_{\mathrm{s},5} \approx \frac{1}{\omega_{\mathrm{n}}}\begin{cases} -0.6\zeta^2 + 8.5\zeta - 3.1 & \text{if } 1 \leq \zeta < 2 \\ 6\zeta - 0.5 & \text{if } \zeta \geq 2 \end{cases}$$

$$t_{\mathrm{s},2} \approx \frac{1}{\omega_{\mathrm{n}}}\begin{cases} -1.1\zeta^2 + 12.3\zeta - 5.3 & \text{if } 1 \leq \zeta < 2 \\ 7.9\zeta - 0.9 & \text{if } \zeta \geq 2 \end{cases}$$

with OS $= 0\%$. Note that all time characteristics inversely proportional to ω_{n}, meaning that

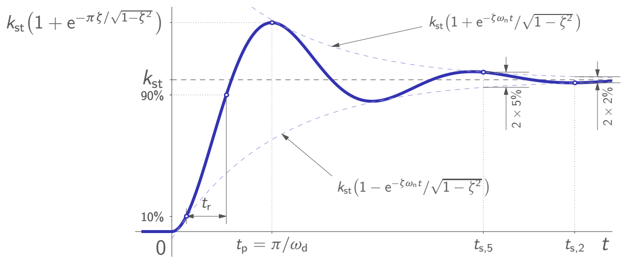
- transients become faster as ω_{n} grows.

2nd order underdamped systems

Step response, where $\omega_{\mathrm{d}} := \omega_{\mathrm{n}}\sqrt{1-\zeta^2}$ is the damped natural frequency,

$$y(t) = k_{\mathrm{st}}\left(1 - \frac{1}{\sqrt{1-\zeta^2}}\,\mathrm{e}^{-\zeta\omega_{\mathrm{n}} t}\sin(\omega_{\mathrm{d}} t + \arccos\zeta)\right)\mathbb{1}(t)$$

or[3]



$$t_{\mathrm{r}} \approx \frac{1.6\zeta^3 - 0.17\zeta^2 + 0.92\zeta + 1.02}{\omega_{\mathrm{n}}}$$

$$t_{\mathrm{s},5} \approx \frac{2.88}{(-\zeta^2 + 0.64\zeta + 0.96)\,\zeta\omega_{\mathrm{n}}}$$

$$t_{\mathrm{s},2} \approx \frac{6.49}{(-\zeta^2 + 0.45\zeta + 1.65)\,\zeta\omega_{\mathrm{n}}}$$

with OS $= \mathrm{e}^{-\pi\zeta/\sqrt{1-\zeta^2}} \cdot 100\%$ (depends only on ζ) and time characteristics again inversely proportional to ω_{n}.

[3] Simpler estimates $t_{\mathrm{s},5} \approx \frac{3}{\zeta\omega_{\mathrm{n}}}$ and $t_{\mathrm{s},2} \approx \frac{3.9}{\zeta\omega_{\mathrm{n}}}$ may be used if $\zeta \ll 1$; however, as $\zeta \uparrow 1$ they might fail by a factor of ≈ 1.5.

Important points

Static gain k_{st} has

- no effect on transients

in both 1st and 2nd order systems, it merely scales the y-axis.

Time constant τ & natural frequency ω_{n} affect transients only through

- scaling the time axis

and do not affect the shape of transient response (smaller τ / larger ω_{n} yield faster response).

Damping factor ζ affects both

- shape of transients

and

- speed of transients

(as ζ decreases, transients become faster albeit more oscillatory).